A Comparison of Price Imputation Methods under Large Samples and Different Levels of Censoring.

Jose A. Lopez
Department of Agricultural Sciences
Texas A&M University – Commerce
Contact: Jose_Lopez@tamu-commerce.edu or lope748@yahoo.com

Abstract

Several imputation approaches under a large sample and different levels of censoring are compared and contrasted by using a multiple imputation methodology. The study not only discusses imputation approaches, but also quantifies differences in price variability before and after price imputation, evaluates the performance of each method, and estimates and compares parameters from a complete demand system. The study's findings reveal that even when there is small variability among different price imputation approaches, there may be large variability among the underlying parameter estimates of interest and the ultimately desired measures. This suggests that a multiple imputation approach may be preferred over a comparison of mean prices among various imputation approaches.

Keywords: imputation methods, multiple imputation, censored prices, protein demand, elasticities

JEL codes: C81, Q11, R21

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2011 AAEA & NAREA Joint Annual Meeting, Pittsburgh, Pennsylvania, July 24-26, 2011

Introduction

Survey design, implementation, and institutional constrains often lead to a frequently encountered problem with consumer survey data, the existence of censored observations. This problem of increasing importance is very common and usually takes place in high proportions (e.g., Taylor et al. 2008, Dong et al. 2004, Gould et al. 2002, Golan et al. 2001, Sabates et al. 2001, Dong and Gould 2000, Heien et al. 1989, Cox and Wohlgenant 1986) in dependent variables, independent variables, or both. It occurs when the value of an observation is partially known (also called item nonresponse). This happens when the value of a variable of interest (e.g., the dependent variable) is unknown; but information on related variables (e.g., the independent variables) is known.

Agricultural economists are very familiar with the use of parametric models when there is item nonresponse on the dependent variable (e.g., the probit and tobit models, or their multinomial versions). However, when there is item nonresponse on the independent variable, a surprisingly high number of studies such as Golan et al. 2001 and Dong et al. 2004 use very simple techniques (e.g., simple regional or quarterly averages) or omit missing observations without being aware of alternative procedures and the key issues involved. There are several ways in which a missing value can be substituted with a replacement value: deductive imputation, cell mean imputation, hot-deck imputation, and cold-deck imputation. Unfortunately, some of these methods may be time consuming (e.g., deductive imputation) or perhaps unfeasible (e.g., cold-deck imputation) when the data sample is large and/or the data is limited.

In deductive imputation the researcher deduces the missing value by using logic and the relationships among the variables. For instance, if the geographical location of a household is missing, it can be recovered by using other variables such as the consecutive order of household interviews and the time period when the household was interviewed. If the previously interviewed and the subsequently interviewed household were interviewed during the same week and they both belong to the same city, then the logical imputation for the missing geographical location would be to use the same city.

Cell mean imputation consists of grouping the observations (e.g., households) into classes (e.g., strata and state) and using the non-missing values of the variable of interest (e.g., non-missing prices) to impute the missing values of the variable of interest (e.g., missing prices). Examples of cell mean imputation include Golan et al. (2001, p. 545) and Dong et al. (2004, p. 1099). Clearly, the more specific the classes (e.g., strata and county) are, the more likely the researcher is to obtain an estimate that is closer to the true value. Cell mean imputation is appropriate if the missing values are missing completely at random. The disadvantage of this method is that the variance in the variable of interest decreases.[1] To avoid losing variability in the variable of interest, the researcher may alternatively use the mean and standard deviation from the non-missing values of the variable of interest and generate values for imputation from a normal distribution with this mean and this standard deviation.

Lohr (1999, p. 275) explains that the term *hot deck* dates back to the time computer programs and datasets were punched on cards. The card reader used to warm the data cards being analyzed, so the term *hot deck* was used to refer to the data cards being analyzed. Similar to cell mean imputation, after the observations have been grouped into classes, hot deck imputation uses a non-missing value of the variable of interest to impute the missing values of the variable of interest. The non-missing value may be the previous non-missing value in the class, a non-

[1] For example, using four strata and Mexico's 31 states plus the Federal District produces 128 different values for the missing values. Using two strata and 32 states/locations produces 64 different values.

missing value chosen at random in the class, or the nearest non-missing value in the cell, where the distance may be defined according to some criteria that is based on another variable.

Contrary to hot deck, cold deck imputation uses a dataset other than the dataset being analyzed to impute the missing value. These datasets may be from a previous survey or from another source. Cold deck imputation is common in time series datasets. The researcher sometimes pulls data from different sources to complete a time series for a particular variable of interest on which few information is available.

The various imputation methods can be compared using a multiple imputation methodology (see Lohr 1999, Rubin 1996, and Rubin 1987). In multiple imputations, a missing value is imputed more than once by using different imputation methods. Each imputation method generates a new dataset with non-missing observations. Each dataset is then analyzed as if no imputation had been done. "[T]he different results give the analyst a measure of the additional variance due to imputation" (Lohr 1999, p. 277). Typically, the same model is used to analyze each imputed dataset.

This research paper compares and contrast several price imputation approaches, under large samples and different levels of censoring, by employing a multiple imputation methodology. The general objective is to discuss and compare imputation approaches by using a complete demand system model on the imputed datasets. In particular, an Almost Ideal Demand System (AIDS) that incorporates the restrictions of adding-up, homogeneity, and symmetry is employed. The paper not only discusses the above mentioned price imputation approaches, but also quantifies the differences in price variability before and after price imputation, evaluates the performance of each method under different levels of missing data, and estimates and compares the ultimately desired parameter estimates (i.e., the parameter estimates obtained from the complete demand system) under each imputation procedure. There are essentially no studies in the existing literature that have studied this critical issue in a large cross-sectional sample.

To accomplish the general objective of the study, data on prices of several important protein sources in the Mexican diet (meat, dairy, eggs, oils, fats, tubers, vegetables, legumes, and fruits) are used from the 2008 survey of Mexican household incomes and expenditures (*Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH)).* ENIGH 2008 is a recent and reliable source of information, and it is published by a Mexican governmental institution (*Instituto Nacional de Estadística, Geografía e Informática (INEGI)).* In ENIGH 2008, a total of 29,468 households were interviewed. However, because the collection period from each household is one week and because ENIGH only records observations when the households make a purchase, the prices of several protein sources are censored. A comparison of price imputation methods under different levels of censoring is ideal with this survey because the sample of households is large.

This study's findings reveal that small variability among the price imputation approaches may lead to large variability among the underlying parameter estimates of interest and the ultimately desired measures (e.g., measures of price responsiveness). This means researchers have to be very careful when choosing a price imputation procedure. In particular, parameter estimates of interest (e.g., the parameter estimates obtained from a demand system) may suffer from selection bias if an imputation method is inappropriately chosen. A simple cell mean price imputation using two levels of urbanization within Mexico's 31 states and the Federal District will produce only 64 different values for price imputation (e.g., Golan et al. 2001; Dong et al. 2004) and may result in a considerable lost of price variability, but may or may not lead to parameter estimates that are not too distant from the true parameters. Similarly, excluding the censored observations may result in parameter estimates that are closer than other approaches to the true parameters.

Price Imputation Methods

The cell mean imputation method is also referred to as a zero-order missing price procedure (Cox and Wolgenant 1986, p. 913). Researches such as Golan et al. (2001, p. 545) and Dong et al. (2004, p. 1099) have used this method. For instance, to impute prices for Mexican households that did not make meat purchases, Golan et al. (2001, p. 545) "assume[d] that those households face the average price level for that product in that particular location: a rural or urban area in a particular state or federal district." Similarly, "[f]or [Mexican] households not purchasing a particular commodity, [Dong et al. (2004, p. 1099)] replace[d] unobserved unit values with the average unit value obtained by purchasing households in the same area, represented by state of residence and degree of urbanization."

Other researchers such as Zheng and Henneberry (2009, p. 878) have used Cox and Wohlgenant's (1986, p. 913) first-order missing price procedure. Using the non-missing prices of commodity i, this method first computes the regional mean prices (mp_i) and then calculates the corresponding deviations from the regional mean prices (dmp_i). That is,

(1) $$dpm_i = p_i - mp_i.$$

Subsequently, this method regresses dmp_i as a function household characteristics, which are proxies for household preferences for unobserved household characteristics. That is,

(2) $$dpm_i = \mathbf{x_i}'\boldsymbol{\beta}_\mathbf{i} + e_i,$$

where $\mathbf{x_i}'$ is a $(1 \times K)$ vector of household characteristics, $\boldsymbol{\beta}_\mathbf{i}$ is a $(K \times 1)$ vector of parameters, and e_i is random error. Cox and Wohlgenant (1986, p. 913) assume that the deviations from mean prices reflect quality differences that are induced by household characteristics and nonsystematic supply-related factors. Substituting equation (2) into (1) and solving for p_i gives the price/quality functions. The OLS parameter estimates obtained from equation (2) are used to predict the values of the missing prices. The quality-adjusted missing price estimates or imputed prices are obtained as follows:

(3) $$\tilde{p}_i = \widehat{dmp}_i + mp_i,$$

where $\tilde{p}_i$ is an estimate of p_i for the corresponding missing prices.

Cox and Wohlgenant's (1986) first-order missing price procedure is really a regression imputation approach. Lopez and Malaga (2009) also use a similar but simpler regression imputation approach. Their approach regresses price as a function of household characteristics, income, and regional dummy variables, and then uses the resulting parameters estimates to predict the missing prices. Other alternative imputation methods include Heckman (1974) and the revised EM-algorithm of Cameron and Trivedi (2005).

Theoretical Model

In multiple imputations, a missing value is imputed more than once. If *m* different imputation methods are used, then *m* different datasets with non-missing values are created. "Each of the *m* data sets is analyzed as if no imputation had been done; the different results give the analyst a measure of the additional variance due to imputation" (Lohr 1999, p. 277).

This paper will compare and contrast different imputation methods under different levels of censoring by estimating an Almost Ideal Demand System (AIDS) that incorporates the restrictions of adding-up, homogeneity, and symmetry for each of the imputed datasets.

The Almost Ideal Demand System (AIDS) was developed by Deaton and Muelbauer (1980) as an arbitrary first order approximation of any demand system. It satisfies the axioms of choice exactly and aggregates perfectly over consumers up to a market demand function without invoking parallel linear Engel curves. The functional form is consistent with household-budget data, can be used to test the properties of homogeneity and symmetry through linear restrictions on fixed parameters, and is not difficult to estimate. In the AIDS model, the Marshallian demand function for commodity *i* in share form is specified as:

$$(4) \qquad w_{ih} = \alpha_i + \sum_j \gamma_{ij} \log(p_{jh}) + \beta_i \log[X_h/P_h] + \varepsilon_{ih},$$

where w_{ih} is the budget share for commodity *i* and household *h*; p_{jh} is the price of commodity *j* and household *h*, X_h is total household expenditure on the commodities being analyzed; α_i, β_i and γ_{ij} are parameters, and ε_i is a random term of disturbances, and P_h is a price index.

In a nonlinear approximation, the price index P_h is defined as:

$$(5) \qquad \log (P_h) = \alpha_0 + \sum_k \alpha_k \log (p_{kh}) + \frac{1}{2} \sum_k \sum_j \gamma_{kj} \log(p_{kh}) \log(p_{jh}).$$

In the linear approximation of the AIDS model (LA/AIDS) suggested by Stone (1954), equation (5) is estimated by:

$$(6) \qquad \log (P^*) = \sum_k w_{kh} \log (p_{kh}).$$

The demand theory properties of adding-up, homogeneity and symmetry can be imposed on the system of equations by restricting parameters in the model as follows:

(7) Adding-up: $\sum_i \alpha_i = 1, \sum_j \gamma_{ij} = 0, \text{ and } \sum_i \beta_i = 0;$

(8) Homogeneity: $\sum_i \gamma_{ij} = 0;$

(9) Symmetry: $\gamma_{ij} = \gamma_{ji}.$

The Marshallian (uncompensated) and the Hicksian (compensated) price elasticities as well as the expenditure elasticities can be computed from the estimated coefficients as follows:

(10) Marshallian Price Elasticity: $e_{ij} = -\delta_{ij} + \gamma_{ij}/w_i - \beta_i\, w_j/w_i$

(11) Hicksian Price Elasticity: $e_{ij}^{c} = -\delta_{ij} + w_j + \gamma_{ij}/w_i$

(12) Expenditure Elasticity: $e_i = 1 + \beta_i /w_i$

where δ is the Kronecker delta equal to one if i = j and equal to zero otherwise.

One equation is omitted in the estimation of this system, but the parameters of that equation will be recovered by making use of the theoretical classical properties. Usually the equation excluded is the one holding the smallest budget share.

Data and Procedures

Mexican data on household income and weekly expenditures was obtained from *Encuesta Nacional de Ingresos y Gastos de los Hogares* (2008), which is a nation-wide survey encompassing Mexico's 31 states plus one Federal District (a territory which belongs to all states). ENIGH is a cross-sectional data sample published since 1977 (e.g., see Heien et al. 1989) by a Mexican governmental institution (*Instituto Nacional de Estadística, Geografía e Informática* (INEGI)). ENIGH collects data by giving direct interviews and recording household expenditures on groceries and several other items for one week.

Seven food sources of protein were analyzed in this study. These are meat (which includes beef, pork, processed meat, chicken, processed poultry meat, seafood, and other meats), dairy (which includes milk, cheese, and other milk derived products), eggs, tubers (which includes raw, fresh, and processed tubers), vegetables (which includes fresh and processed vegetables and pod vegetables), legumes (fresh and processed), and fruits (fresh and processed). More specific information about the food products included in each category can be obtained from ENIGH (2008).

In this study, a subsample of 3,572 households containing non-missing prices and quantities of several important protein sources in the Mexican diet is used. To accomplish the objectives of the study, prices from this dataset were intentionally censored at two levels: a 30% censoring level (3,572 non-missing price observations) and a 70% censoring level (1,702 non-missing price observations). The censoring was done by randomly selecting prices and making them missing prices, and then merging these missing prices with the remaining non-missing prices. The resulting datasets have the 3,572 observations and consists of missing and non-missing prices as indicated by the corresponding percentage (30% censoring or 70% censoring).

Results and Discussion

Table 1 reports the means and standard errors of the various protein categories under no censoring, a 30% censoring level, and a 70% censoring level. The column named No Censoring reports the means and standard errors with non-missing prices only. Method 1 is a cell mean imputation procedure, Method 2 is a simple regression imputation approach, and Method 3 is Cox and Wohlgenant's first-order missing price procedure. These three methods impute the number of missing-price values that is indicated by their column headings. The columns that exclude censored observations (i.e., the columns named Excluding Cen. Obs.) are the only columns that do not report the mean and standard error for the 3,572 households; all other columns use a sample size of 3,572 households.

Table 1 shows that the four approaches to censored prices result in mean values with small variability but standard errors with relatively larger variability. Compared to the dataset with no missing price observations (i.e., the No Censoring column), at a 30% censoring level, the mean prices from the different methods ranged from being 1.02% lower (i.e., the legumes price for methods 2 and 3) to 3.25% higher (i.e., the tubers price for methods 2 and 3). Similarly, at a 30% censoring level, the standard errors of the means ranged from being 37.16% lower (i.e., the legumes price for method 1).

Consider only the different imputation approaches at 30% censoring level, the difference among the means is very small, less than 1%. For the standard errors of the means at a 30% censoring level, the difference among the four imputation approaches is large, on average 30.43% higher or lower. Interestingly among the different imputation approaches, small variability in the means is observed.[2]

At the 70% censoring level, variability increases in both means and standard error of means. Compared to the dataset with no missing price observations, at a 70% censoring level, the mean prices from the different methods ranges from being 6.75% lower (i.e., the tubers price from methods 2 and 3) to 2.24% higher (i.e., the legumes price when excluding the censored observations). The standard errors of the means exhibit an even large variability than the ones at a 30% censoring level. For instance, the standard error of the mean for the meat price from method 1 is 46.92% lower than the same standard error of the mean from the dataset with no missing price observations. Likewise, the standard error of the mean for the legumes price from excluding the censored observations is 147.59% higher than the same standard error of the mean from the dataset with no missing observations.

[2] This does not necessarily mean that there would not be small variability in the ultimately desire measures (i.e., the elasticity measures).

Among the different imputation approaches at 70% censoring level, the difference among the means is also small, less than 1%. For the standard errors of the means at a 70% censoring level, the difference among the three imputation approaches is large, on average 66.8% higher or lower.

The parameter estimates from the simple regression imputation approach (i.e., method 3) at 30% censoring level are reported in Table 2. The variable *p00_11* is the number of household members who are less than 12 years old; *p12_64* is the number of household members who are or are between 12 and 64 years old; *p65_more* is the number of household members who are or are older than 65 years old; *inc* is household income; the variable *rural* takes the value of 1 for household locations with a population of 14,999 people or less and 0 if otherwise; *element* takes the value of 1 if the household decision maker has elementary school education or less and 0 if otherwise; *highsch* takes the value of 1 if the household decision maker has high school education or if he/she is a high school graduate and 0 if otherwise; *college* if the household decision maker has some college, college or incomplete university education and 0 if otherwise; *NE, NW, CW,* and *SE* are regional dummy variables for the Northeast, Northwest, Central-West, and Southeast regions respectively, *d_car* and *d_refri* are dummy variables for car and refrigerator respectively, and *supermkt* takes the value of 1 if the household purchased the protein product or commodity from a supermarket and 0 if somewhere else. The excluded dummy variables are *urban, university,* and *C*. This means that the baseline consists of urban households from the central region whose decision makers have completed university education or graduate school education.

The parameter estimates reported in Table 2 were used to impute missing price observations at the corresponding 30% censoring level. The resulting mean and average standard error of the mean under method 3 are reported in Table 1 under the 30% censoring section. From a total of 16 parameter estimates, on average, about 5 parameter estimates were statistically different from zero at the 0.05 significance level and about 7 parameter estimates at the 0.20 significance level.[3] Overall, the simple regression imputation approach (method 3) had a slightly higher total number of parameters statistically different from zero when compared to the first-order missing price procedure. Given that there were small difference among these two regression imputation approaches, and the fact that method 3 is a simpler approach to method 1, only the results from method 3 are presented and discussed from here on.

Table 3 reports the parameter estimates from full AIDS models, equations (4) and (5), estimated under the various approaches to price imputation (i.e., excluding the censoring observations, using a cell mean imputation approach, or using a simple regression imputation approach) for a 30% censoring level. From a total of 41 parameters estimated, at least 32 are statistically different from zero at the 0.05 significance level for each approach. Compared to the parameter estimates obtained from the dataset with no censored prices, the parameter estimates from the different approaches are on average 33% higher or lower. The difference ranged from being 414.47% lower (i.e., g27 from method 3) to 173.65% higher (i.e., g35 from the approach that excludes the censored observations). These differences are remarkably higher when using the datasets with 70% of the observations censored.[4] Surprisingly, the parameter estimates among the various censoring approaches that were on average closer to the parameter estimates under no censoring are the ones that correspond to excluding the censored observations.

[3] The parameter estimates from the simple regression imputation approach (i.e., method 3) at a 70% censoring level and the parameter estimates from the first-order missing price procedure of Cox and Wohlgenant (1986) (i.e., method 2) under both the 30% censoring level and the 70% censoring level are available upon request.
[4] The AIDS parameter estimates under the various approaches to price imputation for a 70% censoring level are available upon request.

Tables 4, 5, and 6 show the Marshallian own-price elasticities, the Hicksian own-price elasticities, and the expenditure elasticities respectively. Differences are also observed between the different censoring approaches and the elasticities obtained from the dataset with no censored observations. Compared to the no-cenrored Marshallian own-price elasticities, the elasticities from the different approaches are on average 6.85% higher or lower (Table 4). Compared to the no-censored Hicksian own-price elasticities, the elasticities from the different approaches are on average 7.72% higher or lower (Table 5). Similarly, compared to the no-censored Expenditure elasticities, the elasticities from the different approach are on average 3.25% higher or lower (Table 6). They range from 48.21% lower ($\hat{e}_{44}$, method 3, and 70% censoring level) to 10.10% higher ($\hat{e}_{66}$, excluding censored observations, and 70% censoring level), from 50.53% lower ($\hat{e}_{44}^{c}$, method 3, and 70% censoring level) to 10.55% higher ($\hat{e}_{66}^{c}$, excluding censoring observations, and 70% censoring level), and from 10.21% higher ($\hat{e}_{6}$, method 3, and 70% censoring level) to 12% higher ($\hat{e}_{2}$, method 1, and 70% censoring level) respectively. Consistent with the results from the AIDS parameter estimates, the approach of excluding the censored observations was closer to the elasticity estimates obtained under no censoring.

Considering only the elasticities obtained from the various imputation approaches, some considerable differences are observed, especially at the 70% censoring level. For instance, the Marshallian own-price elasticities are on average 10.30% lower or higher than the same elasticities obtained when excluding the censored observations. Likewise, the hicksian own-price elasticities are on average 11.55% lower or higher and the expenditure elasticities are on average 5.23% lower or higher.

Interestingly, this means that even when there was small variability in the mean prices among the various price imputation approaches (Table 1), considerable larger variability was found in the ultimately desired elasticity measures among the various price imputation approaches (Tables 4, 5, and 6). This suggests that a multiple imputation approach may be preferred over a comparison of mean prices among various imputation approaches.

Concluding Remarks

Several studies use simple techniques to account for the problem of censored prices. These approaches either omit the missing prices or use price imputation approaches such as deductive imputation, cell mean imputation, hot-deck imputation, cold-deck imputation, and regression imputation. This study compares and contrast four imputation approaches (excluding censoring observations, a cell mean imputation approach, the first-order missing price procedure of Cox and Wolhgenant (1986), and a simple regression imputation approach) under two levels of censoring by using a multiple imputation methodology. The study also quantifies the differences in price variability before and after price imputation, evaluates the performance of each method under different levels of missing data, and estimates and compares the ultimately desired parameter estimates (i.e., the parameter estimates obtained from the complete demand system) under each imputation procedure.

The study's findings reveal that even when there is small variability among different price imputation approaches there may be large variability among the underlying parameter estimates of interest as well as the ultimately desired measures (e.g., measures of price responsiveness). In particular, parameter estimates may suffer from selection bias if an imputation method is inappropriately chosen. A simple cell mean imputation method which may result in a considerable lost of price variability may also allow for parameter estimates that are not too distance from the true parameters. Similarly, excluding censored observations may result in parameter estimates that are closer to the true parameters.

This study not only increases awareness of different approaches to price imputation but also illustrates that the models that are ultimately estimated (such as the AIDS model) are sometimes sensitive to the price imputation approach taken. Further research with additional imputation methods and different types of samples (e.g., time series data or panel data) would complement the study. Research on datasets with prices not missing at random also needs to be explored. In addition, further research is needed on determining more specific conditions under which excluding observations would be preferred over imputation approaches.

Finally, this study computed Marshallian and Hicksian own- and cross-price elasticities as well as expenditure elasticities for several important protein sources (meat, dairy, eggs, tubers, vegetables, legumes, and fruits) in the Mexican diet. The elasticity estimates are relatively recent and contribute to a better understanding of the Mexican demand for sources of protein. The elasticity estimates can also be employed to analyze current and/or future trends in protein consumption.

References

Cameron, A.C., and P.K. Trivedi. Microeconometrics: Methods and Applications. Cambridge: Cambridge University Press, 2005.

Cox, T.L., and M.K. Wohlgenant. "Prices and Quality Effects in Cross-Sectional Demand Analysis." American Journal of Agricultural Economics 68(1986):908–919.

Deaton, A., and Muellbauer. "An Almost Ideal Demand System." The American Economic Review 70, 3 (1980): 312-325.

Dong, D., and B.W. Gould. "Quality Versus Quantity in Mexican Poultry and Pork Purchases." Agribusiness 16(Summer 2000):333–355.

Dong, D., B.W. Gould, and H.M. Kaiser. "Food Demand in Mexico: An Application of the Amemiya-Tobin Approach to the Estimation of a Censored Food System." American Journal of Agricultural Economics 86(November 2004):1094–1107.

Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH). Aguascalientes, México: Instituto Nacional de Estadística, Geografía e Informática (INEGI), 2008. Internet site: http://www.inegi.gob.mx.

Gould, B.W., Y. Lee, D. Dong, and H.J. Villarreal. "Household Size and Composition Impacts on Meat Demand in Mexico: A Censored Demand System Approach." Paper presented at the American Agricultural Economics Association Annual Meeting, Long Beach, California, July 28–31, 2002.

Golan, A., J.M. Perloff, and E.Z. Shen. "Estimating a Demand System with Nonegativity Constraints: Mexican Meat Demand." The Review of Economics and Statistics 83(August 2001):541–550.

Heckman, J.J. "Shadow Prices, Market Wages, and Labor Supply." Econometrica 42(1974):679-694.

Heien, D., L.S. Jarvis, and F. Perali. "Food Consumption in Mexico: Demographic and Economic Effects." Food Policy Journal 14(May 1989):167–179.

Lohr, S.L. Sampling: Design and Analysis. New York: Duxbury Press, 1999.

Lopez, J.A. and J.E. Malaga. "Estimation of a Censored Demand System in Stratified Sampling: An Analysis of Mexican Meat Demand at the Table Cut Level." Paper presented at the Southern Agricultural Economics Association Annual Meeting, Atlanta, Georgia, January 31 – February 3, 2009.

Taylor, M., D. Phaneuf, and N. Piggott. "Does Food Safety Information Affect Consumers' Decision to Purchase Meat and Poultry? Evidence from U.S. Household Level Data." Paper presented at the Western Agricultural Economics Association Annual Meeting, Big Sky, MT, June 23–24, 2008.

Rubin, D.B. Multiple Imputation for Nonresponse in Surveys. New York: Wiley, 1987.

Rubin, D.B. "Multiple Imputation After 18+ Years." Journal of the American Statistical Association 91 (1996):473-489.

Sabates, R., and B.W. Gould, and H. Villarreal. "Household Composition and Food Expenditures: A Cross-Country Comparison." Food Policy 26(July 2001):571–586.

Stone, J.R.N. "The Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand" Economic Journal 64(1954):511-527.

Zheng, Z., and S.R. Henneberry. "An Analysis of Food Demand in China: A Case Study of Urban Households in Jiangsu Province." Review of Agricultural Economics 31(2009):873-893.

Table 1. Observed and Imputed Prices (n = 3,572).

	No Censoring		30 % Censoring Level							
	Observed Prices		Excluding Cen. Obs.		Method 1		Method 2		Method 3	
p_i	Mean (Pesos/Kg)	Std. Err. of Mean	Mean (Pesos/Kg)	Std. Err. of Mean	Mean (Pesos/Kg)	Std. Err. of Mean	Mean (Pesos/Kg)	Std. Err. of Mean	Mean (Pesos/Kg)	Std. Err. of Mean
p1	46.4608	0.3650	47.0064	0.4462	47.0064	0.3071	46.9953	0.3124	46.9953	0.3124
p2	23.7807	0.4708	23.9239	0.5504	23.9239	0.3785	23.8325	0.3874	23.8325	0.3874
p3	18.7620	0.1311	18.8758	0.1769	18.8758	0.1216	18.8804	0.1242	18.8804	0.1242
p4	15.5820	0.5964	16.0031	0.7511	16.0031	0.5165	16.0884	0.5180	16.0884	0.5180
p5	13.3280	0.1362	13.1985	0.1662	13.1985	0.1143	13.2155	0.1173	13.2155	0.1173
p6	18.6618	0.2500	18.4720	0.2282	18.4720	0.1571	18.5022	0.1591	18.5022	0.1591
p7	10.3969	0.1455	10.4638	0.1685	10.4638	0.1159	10.4776	0.1177	10.4776	0.1177

	No Censoring		70 % Censoring Level							
	Observed Prices		Excluding Cen. Obs.		Method 1		Method 2		Method 3	
p_i	Mean (Pesos/Kg)	Std. Err. of Mean	Mean (Pesos/Kg)	Std. Err. of Mean	Mean (Pesos/Kg)	Std. Err. of Mean	Mean (Pesos/Kg)	Std. Err. of Mean	Mean (Pesos/Kg)	Std. Err. of Mean
p1	46.4608	0.3650	45.2598	0.6193	45.2598	0.1938	45.3696	0.2156	45.3696	0.2156
p2	23.7807	0.4708	23.4655	0.8953	23.4655	0.2794	23.6935	0.3108	23.6935	0.3108
p3	18.7620	0.1311	18.5115	0.1558	18.5115	0.0487	18.4960	0.0547	18.4960	0.0547
p4	15.5820	0.5964	14.6550	0.9537	14.6550	0.2977	14.5298	0.3079	14.5298	0.3079
p5	13.3280	0.1362	13.6131	0.2372	13.6131	0.0740	13.6234	0.0834	13.6234	0.0834
p6	18.6618	0.2500	19.0796	0.6189	19.0796	0.1937	19.0082	0.2119	19.0082	0.2119
p7	10.3969	0.1455	10.2498	0.2817	10.2498	0.0879	10.2020	0.0926	10.2020	0.0926

Note: p_i, $i = 1, \ldots, 5$, where 1 = meat, 2 = dairy, 3 = eggs, 4 = tubers, 5 = vegetables, 6 = legumes, and 7 = fruits. Average exchange rate in 2008 is US $1 = 11.14 Pesos (Banco de México).

Source: ENIGH 2008 Database, computed by authors.

Table 2. Parameter Estimates from the Simple Imputation Approach (Method 3) and 30% Censoring Level.

Var.	Meat			Dairy			Eggs			Tubers			Vegetables			Legumes			Fruits		
	Param. Est.		Std. Error	Param. Est.		Std. Error	Param. Est.		Std. Error	Param. Est.		Std. Error	Param. Est.		Std. Error	Param. Est.		Std. Error	Param. Est.		Std. Error
Intercept	1.2832		2.8392	3.2118		2.5888	-0.3437		0.8028	4.3904		3.8939	-1.2768	***	0.5346	-0.1709		1.0740	0.5266		0.7589
p00_11	-1.4781	***	0.3542	-0.7985	***	0.4059	-0.0014		0.1010	0.5988		0.6251	-0.0160		0.0846	-0.0798		0.1526	-0.1486		0.1755
p12_64	-0.2838		0.2773	0.2240		0.3493	-0.0151		0.0729	-0.6717	**	0.3950	-0.0923	*	0.0580	-0.2741	***	0.1017	-0.3222	***	0.1134
p65_more	0.6273		0.9124	-0.4737		0.9955	0.5614	*	0.3628	-2.3071	***	1.1596	-0.3504	***	0.1575	-0.0912		0.4091	0.0913		0.3782
inc	0.0000	***	0.0000	0.0000	*	0.0000	0.0000		0.0000	0.0000		0.0000	0.0000	***	0.0000	0.0000	*	0.0000	0.0000	***	0.0000
rural	-0.9933		1.0168	-1.4153		1.2388	0.1652		0.3232	-1.7267		1.7417	0.6464	***	0.3149	-1.7093	***	0.5081	-0.0728		0.3820
element	-4.4806	***	2.0119	1.9825		1.8370	-0.6132		0.7720	-3.4712		3.5912	-0.7953	**	0.4128	-0.6083		0.9283	-0.6967		0.6001
highsch	-3.7244	**	2.0824	0.8528		1.8378	-0.5106		0.7453	-3.7848		3.6786	-0.2370		0.4397	-0.6865		0.9314	-1.4103	***	0.6101
college	-1.1876		2.1631	1.8367		2.6946	1.0564		1.3744	-3.0762		3.8396	-0.6058	*	0.4250	0.6364		0.9790	-1.4283	***	0.5998
NE	-1.0669		1.6261	-4.3916	***	1.4554	3.7242	***	0.8593	5.0986	**	2.6716	3.0612	***	0.3602	3.8294	***	0.9123	2.7245	***	0.7140
NW	1.7696		2.3315	-5.7355	***	1.6467	1.0347	*	0.7577	3.9231	*	2.4857	3.4372	***	0.5663	-0.5041		0.9535	1.4864	***	0.6981
CW	3.0536	***	1.0579	-5.4597	***	1.1395	-0.1827		0.2358	1.9255		1.8466	0.6768	***	0.2925	2.7562	***	0.4821	1.4975	***	0.3797
SE	0.9414		1.2407	10.1579	***	2.1295	2.6736	***	0.7642	2.9304	*	1.9253	2.1033	***	0.2838	-0.2382		0.8493	3.0419	***	0.5850
d_car	1.5631	*	1.0007	-0.6302		1.2167	-0.3489		0.4246	-1.5354		1.6195	-0.0666		0.2402	-0.1886		0.5534	0.1627		0.3579
d_refri	1.8758	*	1.2166	-3.0685	**	1.8235	-0.5315		0.4991	-0.1922		2.4715	0.5362	**	0.3170	1.1250	***	0.5731	-0.2909		0.6577
supermkt	-1.8473		1.6040	0.3380		2.1925	1.7374	*	1.2624	-1.1470		1.9952	0.9228	***	0.3945	-0.1214		0.8047	0.9652	**	0.5460
R-square	0.0706			0.0849			0.0581			0.0087			0.1017			0.0508			0.0516		
F-Value	12.58	***		15.35	***		10.22	***		1.46	*		26.83	***		8.86	***		9.02	***	

Note: Significance levels of 0.05, 0.10, and 0.20 are indicated by triple asterisks (***), double asterisks (**), and an asterisk (*) respectively.

Table 3. AIDS Parameter Estimates Under 30% Censoring.

				30% Censoring								
	No Censoring			Excluding			Method 1			Method 3		
Par.	Estimate		Approx Std Err	Estimate		Approx Std Err	Estimate		Approx Std Err	Estimate		Approx Std Err
g11	0.0269	***	0.0062	0.0283	***	0.0075	0.0324	***	0.0074	0.0303	***	0.0073
g12	0.0203	***	0.0031	0.0203	***	0.0037	0.0146	***	0.0038	0.0228	***	0.0037
g13	-0.0261	***	0.0021	-0.0292	***	0.0025	-0.0292	***	0.0025	-0.0310	***	0.0024
g14	-0.0037	***	0.0012	-0.0047	***	0.0014	-0.0046	***	0.0016	-0.0064	***	0.0016
g15	-0.0059	**	0.0032	-0.0012		0.0039	-0.0001		0.0038	0.0003		0.0038
g16	-0.0148	***	0.0022	-0.0172	***	0.0027	-0.0169	***	0.0026	-0.0182	***	0.0026
g17	0.0033	*	0.0025	0.0038	*	0.0030	0.0038		0.0030	0.0022		0.0030
g22	-0.0199	***	0.0029	-0.0238	***	0.0034	-0.0160	***	0.0038	-0.0225	***	0.0035
g23	-0.0037	***	0.0012	-0.0034	***	0.0014	-0.0041	***	0.0014	-0.0041	***	0.0013
g24	-0.0019	***	0.0007	-0.0017	***	0.0008	-0.0003		0.0009	-0.0016	**	0.0009
g25	0.0082	***	0.0019	0.0104	***	0.0022	0.0085	***	0.0023	0.0089	***	0.0022
g26	-0.0031	***	0.0012	-0.0020	*	0.0015	-0.0026	**	0.0014	-0.0029	***	0.0014
g27	0.0002		0.0015	0.0003		0.0017	0.0000		0.0018	-0.0005		0.0018
g33	0.0264	***	0.0023	0.0248	***	0.0027	0.0260	***	0.0026	0.0272	***	0.0026
g34	0.0039	***	0.0009	0.0042	***	0.0011	0.0024	***	0.0011	0.0028	***	0.0011
g35	0.0010		0.0019	0.0017		0.0022	0.0027		0.0022	0.0027		0.0022
g36	0.0027	**	0.0015	0.0045	***	0.0018	0.0047	***	0.0017	0.0047	***	0.0017
g37	-0.0042	***	0.0014	-0.0026	*	0.0016	-0.0027	**	0.0016	-0.0024	*	0.0016
g44	0.0072	***	0.0007	0.0067	***	0.0008	0.0087	***	0.0009	0.0101	***	0.0009
g45	-0.0032	***	0.0011	-0.0025	***	0.0012	-0.0041	***	0.0013	-0.0036	***	0.0013
g46	-0.0012	*	0.0008	-0.0008		0.0010	-0.0013		0.0010	-0.0008		0.0010
g47	-0.0010	*	0.0008	-0.0011		0.0009	-0.0008		0.0010	-0.0005		0.0010
g55	0.0181	***	0.0033	0.0129	***	0.0039	0.0140	***	0.0039	0.0129	***	0.0039
g56	-0.0058	***	0.0018	-0.0078	***	0.0021	-0.0076	***	0.0021	-0.0070	***	0.0021
g57	-0.0125	***	0.0018	-0.0134	***	0.0022	-0.0134	***	0.0022	-0.0142	***	0.0022
g66	0.0250	***	0.0019	0.0267	***	0.0023	0.0270	***	0.0022	0.0268	***	0.0022
g67	-0.0028	***	0.0013	-0.0034	***	0.0016	-0.0033	***	0.0016	-0.0026	**	0.0016
g77	0.0170	***	0.0020	0.0163	***	0.0023	0.0163	***	0.0024	0.0179	***	0.0024
a1	0.2673	***	0.0097	0.2689	***	0.0116	0.2754	***	0.0109	0.2662	***	0.0107
a2	0.1377	***	0.0064	0.1372	***	0.0076	0.1247	***	0.0075	0.1292	***	0.0072
a3	0.1506	***	0.0033	0.1545	***	0.0040	0.1536	***	0.0036	0.1557	***	0.0035
a4	0.0641	***	0.0019	0.0648	***	0.0023	0.0687	***	0.0024	0.0714	***	0.0024
a5	0.1896	***	0.0055	0.1823	***	0.0065	0.1835	***	0.0061	0.1814	***	0.0060
a6	0.1301	***	0.0035	0.1345	***	0.0043	0.1321	***	0.0039	0.1333	***	0.0038
a7	0.0606	***	0.0044	0.0579	***	0.0051	0.0620	***	0.0050	0.0628	***	0.0050
b1	0.0447	***	0.0046	0.0452	***	0.0055	0.0324	***	0.0048	0.0395	***	0.0047
b2	0.0312	***	0.0037	0.0312	***	0.0044	0.0523	***	0.0042	0.0408	***	0.0040

Table 3. *Continued*.

	No Censoring			30% Censoring: Excluding			30% Censoring: Method 1			30% Censoring: Method 3		
Par.	Estimate		Approx Std Err	Estimate		Approx Std Err	Estimate		Approx Std Err	Estimate		Approx Std Err
b3	-0.0345	***	0.0015	-0.0341	***	0.0018	-0.0352	***	0.0015	-0.0349	***	0.0015
b4	-0.0133	***	0.0009	-0.0133	***	0.0011	-0.0141	***	0.0010	-0.0137	***	0.0010
b5	-0.0133	***	0.0026	-0.0133	***	0.0031	-0.0173	***	0.0026	-0.0162	***	0.0026
b6	-0.0335	***	0.0016	-0.0350	***	0.0020	-0.0348	***	0.0016	-0.0339	***	0.0016
	R-sqr						R-sqr			R-sqr		
w1	0.0384						0.0221			0.0339		
w2	0.0381						0.0499			0.0446		
w3	0.1780						0.1636			0.1870		
w4	0.0872						0.0760			0.0778		
w5	0.0265						0.0245			0.0245		
w6	0.1430						0.1489			0.1434		

Note: Significance levels of 0.05, 0.10, and 0.20 are indicated by triple asterisks (***), double asterisks (**), and an asterisk (*) respectively.

Table 4. Marshallian Own-Price Elasticities.

Table entries estimate e_{ij}.

i = j	30% Exclude	30% M1	30% M3	70% Exclude	70% M1	70% M3	No Censor	Min.	Max.
1	-0.9219	-0.9085	-0.9148	-0.9367	-0.9015	-0.8964	-0.9253	-0.9367	-0.8964
2	-1.1207	-1.0758	-1.1090	-1.0522	-0.9926	-1.0421	-1.0999	-1.1207	-0.9926
3	-0.6786	-0.6490	-0.6364	-0.5838	-0.4993	-0.4618	-0.6564	-0.6786	-0.4618
4	-0.8312	-0.7960	-0.7661	-0.7816	-0.7480	-0.4245	-0.8197	-0.8312	-0.4245
5	-0.9231	-0.9143	-0.9216	-0.8317	-0.8235	-0.8384	-0.8929	-0.9231	-0.8235
6	-0.6292	-0.6101	-0.6174	-0.7135	-0.6979	-0.5818	-0.6480	-0.7135	-0.5818
7	-0.8062	-0.8088	-0.7916	-0.7796	-0.7850	-0.7872	-0.7997	-0.8088	-0.7796
Min.	-1.1207	-1.0758	-1.1090	-1.0522	-0.9926	-1.0421	-1.0999	-1.1207	
Max.	-0.6292	-0.6101	-0.6174	-0.5838	-0.4993	-0.4245	-0.6480		-0.4245

Note: i, j = 1, 2, …,7, where 1 = meat, 2 = dairy, 3 = eggs, 4 = tubers, 5 = vegetables, 6 = legumes, and 7 = fruits.

Table 5. Hicksian Own-Price Elasticities.

Table entries estimate e_{ij}^{c}.

i = j	30% Exclude	30% M1	30% M3	70% Exclude	70% M1	70% M3	No Censor	Min.	Max.
1	-0.5148	-0.5222	-0.5199	-0.5379	-0.5458	-0.5230	-0.5207	-0.5458	-0.5148
2	-0.8921	-0.8119	-0.8614	-0.8176	-0.6941	-0.7686	-0.8696	-0.8921	-0.6941
3	-0.6354	-0.6099	-0.5966	-0.5429	-0.4620	-0.4231	-0.6139	-0.6354	-0.4231
4	-0.8050	-0.7677	-0.7367	-0.7548	-0.7183	-0.3924	-0.7933	-0.8050	-0.3924
5	-0.7688	-0.7681	-0.7737	-0.6745	-0.6762	-0.6896	-0.7375	-0.7737	-0.6745
6	-0.5921	-0.5756	-0.5814	-0.6750	-0.6654	-0.5480	-0.6106	-0.6750	-0.5480
7	-0.7027	-0.7071	-0.6872	-0.6763	-0.6860	-0.6875	-0.6963	-0.7071	-0.6763
Min.	-0.8921	-0.8119	-0.8614	-0.8176	-0.7183	-0.7686	-0.8696	-0.8921	
Max.	-0.5148	-0.5222	-0.5199	-0.5379	-0.4620	-0.3924	-0.5207		-0.3924

Note: *i*, *j* = 1, 2, …,7, where 1 = meat, 2 = dairy, 3 = eggs, 4 = tubers, 5 = vegetables, 6 = legumes, and 7 = fruits.

Table 6. Expenditure Elasticities

Table entries estimate e_i.

i	30% Exclude	30% M1	30% M3	70% Exclude	70% M1	70% M3	No Censor	Min.	Max.
1	1.1249	1.0914	1.1111	1.1224	1.0548	1.0890	1.1241	1.0548	1.1249
2	1.1583	1.2472	1.1974	1.1546	1.2956	1.2428	1.1568	1.1546	1.2956
3	0.5591	0.5261	0.5323	0.5364	0.5280	0.5380	0.5517	0.5261	0.5591
4	0.6632	0.6688	0.6825	0.6693	0.6699	0.7091	0.6646	0.6632	0.7091
5	0.9207	0.8941	0.9015	0.9186	0.8926	0.8972	0.9214	0.8926	0.9214
6	0.5146	0.4985	0.5156	0.5631	0.4735	0.4884	0.5273	0.4735	0.5631
7	1.2285	1.1960	1.2133	1.2032	1.1814	1.1767	1.2211	1.1767	1.2285
Min.	0.5146	0.4985	0.5156	0.5364	0.4735	0.4884	0.5273	0.4735	
Max.	1.2285	1.2472	1.2133	1.2032	1.2956	1.2428	1.2211		1.2956

Note: *i* = 1, 2, …,7, where 1 = meat, 2 = dairy, 3 = eggs, 4 = tubers, 5 = vegetables, 6 = legumes, and 7 = fruits.